

# The synergy between protein positioning and DNA elasticity: energy minimization of protein-decorated DNA minicircles

Nicolas Clauvelin<sup>a</sup>, Wilma K. Olson<sup>b</sup>

<sup>a</sup>*BioMaPS Institute for Quantitative Biology, Rutgers, the State University of New Jersey, Piscataway, NJ, USA*

<sup>b</sup>*BioMaPS Institute for Quantitative Biology and Department of Chemistry and Chemical Biology, Rutgers, the State University of New Jersey, Piscataway, NJ, USA*

---

## Abstract

The binding of proteins onto DNA contributes to the shaping and packaging of genome as well as to the expression of specific genetic messages. With a view to understanding the interplay between the presence of proteins and the deformation of DNA involved in such processes, we developed a new method to minimize the elastic energy of DNA fragments at the mesoscale level. Our method makes it possible to obtain the optimal pathways of protein-decorated DNA molecules for which the terminal base pairs are spatially constrained. We focus in this work on the deformations induced by selected architectural proteins on circular DNA. We report the energy landscapes of DNA minicircles subjected to different levels of torsional stress and containing one or two proteins as functions of the chain length and spacing between the proteins. Our results reveal cooperation between the elasticity of the double helix and the structural distortions of DNA induced by bound proteins. We find that the imposed mechanical stress influences the placement of proteins on DNA and that, the proteins, in turn, modulate the mechanical stress and thereby broadcast their presence along DNA.

*Keywords:* DNA, elasticity, numerical optimization

---

## 1. Introduction

The organization of long genomes in the confined spaces of a cell requires special facilitating mechanisms. A variety of architectural proteins play key roles in these processes. Some of these proteins help to compact the DNA by introducing sharp turns along its pathway while others bring distant parts of the chain molecule into close proximity. The histone-like HU protein from *Escherichia coli* strain U93 and the structurally related Hbb protein from *Borrelia burgdorferi* induce some of the largest known deformations of DNA double-helical structure—including global bends in excess of 160 deg, appreciable untwisting, and accompanying dislocations of the helical axis [1–5]. The degree of DNA deformation reflects both the phasing of the protein-induced structural distortions and the number of nucleotides wrapped on the protein surface. Changes in the spacing between deformation sites on DNA, one such deformation associated with each half of these two dimeric proteins, contribute to chiral bends, while variation in the electrostatic surface of the proteins perturbs the degree of association with DNA and the extent of duplex bending.

By contrast, recent studies have highlighted the important role played by proteins which bridge and scaffold distant DNA sites and induce the formation of loops. For example, the *Escherichia coli* H-NS (histone-like nucleoid structuring) protein forms a superhelical scaffold [6] to compact DNA and the *Escherichia coli* terminus-containing factor MatP can bridge two DNA sites to form a loop [7]. The MatP protein is responsible for the compaction of a specific macrodomain (Ter) in *Escherichia coli* and is believed to be able to bind sites located on either separate chromosomes or within one chromosome [8]. The detailed

---

*Email address:* clauvelin@biomaps.rutgers.edu (Nicolas Clauvelin)

influence of MatP on chromosome segregation and cellular division is yet to be unraveled [8], but it is expected that the presence of MatP proteins could significantly alter the organization and the dynamics of the whole Ter macrodomain [7].

The composite kinking and wrapping of DNA on the surface of HU resembles, albeit on a much smaller scale, the packaging of DNA around the histone octamer of the nucleosome core particle [9], while MatP might serve as a bacterial analog of the insulator-binding proteins that divide chromatin into independent functional domains [10]. Deciphering how these two very different bacterial proteins influence the structure and properties of DNA provides a first step in understanding how architectural proteins contribute to the spatial organization of genomes.

With this goal in mind, we have recently developed a new method for the optimization of DNA structure at the base-pair level. Our method takes account of the sequence-dependent elasticity of DNA and can be applied to chain fragments in which the first and last base pairs are spatially constrained. Moreover, our approach allows for constraints in intervening parts of the DNA and thus makes it possible to model the presence of bound proteins on DNA. We can accordingly compute the energy landscape for a wide variety of protein-DNA systems and herein illustrate the effects of HU and Hbb on the configurations of circular DNA and provide examples of protein-mediated loops of the type induced by MatP.

The advantage of our method resides in the direct control of the positions and orientations of the base pairs at the ends of a DNA chain and in the capability of accounting for the presence of bound proteins. In addition, the different constraints due to the boundary conditions and the presence of proteins are directly integrated in the minimization process, which makes it possible to rely on unconstrained numerical optimization methods. Others have derived methods to optimize the energy of elastic deformation for DNA fragments. For example, Coleman *et al* [11] developed a method similar to ours, but their approach require explicit specification of the forces and moments acting on the terminal base pairs (that is, there is no direct control on the positions and orientations of the terminal base pairs). Zhang and Crothers also presented an optimization method [12], albeit limited to circular DNA. Their method only considers angular deformations within the double helix and takes the boundary conditions into account through Lagrange multipliers.

We start with a brief description of our method (Section 2) and then present our results for the effect of different proteins on circular DNA. The various appendices contain the details of the calculations.

## 2. Energy minimization of spatially constrained, protein-decorated DNA

We present in this section a novel method for minimizing the elastic energy of a collection of DNA base-pairs. Our method accounts for the sequence-dependent elasticity of DNA and can be applied to a DNA fragment in which the first and last base pairs are spatially constrained. Moreover, our approach makes it possible to constrain parts of the DNA in order to model the presence of bound proteins. We describe our method in a symbolic fashion for readability and present full details of the calculations in the appendices.

### 2.1. Geometry of a collection of base pairs

We consider a collection of  $N$  rigid base pairs and for the  $i$ -th base pair we denote  $\underline{x}^i$  its origin and  $\underline{d}^i$  the matrix containing the axes of the base-pair frame organized as column vectors. The segmented curve defined by the base-pair origins  $\underline{x}^i$  is referred to as the collection centerline and can be interpreted as a discrete double-helical axis.

Here and in the rest of this paper vector symbols are underlined and matrices are represented with bold symbols. Superscripted Roman letters ( $i, j, \dots$ ) are used to index base pairs and base-pair steps.

#### 2.1.1. Base-pair step parameters

The geometry of a collection of base pairs is traditionally described in terms of the base-pair step parameters (or step parameters for short) [13]. These parameters describe the relative arrangement of successive base pairs and are denoted for the  $i$ -th step  $\underline{p}^i = (\theta_1^i, \theta_2^i, \theta_3^i, \rho_1^i, \rho_2^i, \rho_3^i)$  [11] and referred to, respectively, as tilt, roll, twist, shift, slide and rise. The first three parameters correspond to three angles describing the relative orientation of base-pair frames  $\underline{d}^i$  and  $\underline{d}^{i+1}$  and the last three parameters are the

components of the step joining vector  $\underline{r}^i = \underline{x}^{i+1} - \underline{x}^i$  expressed in a particular frame referred to as the step frame (see Appendix A for more details and [11, 14] for explicit calculation methods). The set of step parameters for the base-pair collection is a  $6(N - 1)$  vector and is denoted  $\underline{P}$ .

We shall see later that the energy of elastic deformation for a base-pair step can be written as a quadratic form with respect to the step parameters. The step parameters, however, are not a convenient representation to express constraints on the positions and orientations of base-pairs within the collection. This is due to the fact that the components of the step joining vector  $\underline{r}^i$  depend on the relative orientation of the two base pairs forming the step. In order to circumvent this issue we introduce a different representation of the geometry of a base-pair collection.

### 2.1.2. Base-pair step degrees of freedom

We now define a new set of variables for each step of the base-pair collection. These variables are denoted  $\underline{\phi}^i = (\psi_1^i, \psi_2^i, \psi_3^i, r_1^i, r_2^i, r_3^i)$  for the  $i$ -th step and referred to as the step degrees of freedom or step dofs for short. Although the variables  $\psi^i$  are identical to the angular step parameters  $\theta^i$ , we use a different symbol for clarity. The variables  $r^i$  are the components of the step joining vector  $\underline{r}^i = \underline{x}^{i+1} - \underline{x}^i$  expressed with respect to the global reference frame (a convenient choice for the global reference frame is the first base-pair frame). The set of step dofs for the base-pair collection is a  $6(N - 1)$  vector and is denoted  $\underline{\Phi}$ .

The main advantage of this choice of variables is to separate the representation of the centerline of the base-pair collection from the orientation of the base-pair frames. This is somewhat equivalent to the centerline/spin representation of Langer and Singer [15] for continuous elastic rods. We shall see that this representation is particularly convenient to deal with the end conditions applied to a collection of base pairs.

### 2.2. Energy of elastic deformation

The energy of elastic deformation for a DNA base-pair collection is defined as the sum of the energy of elastic deformation for each step. For the  $i$ -th step, this energy is given by the following quadratic form:

$$E^i = \frac{1}{2} (\underline{p}^i - \underline{\bar{p}}^i)^\top \mathbf{F}^i (\underline{p}^i - \underline{\bar{p}}^i), \quad (1)$$

where  $\underline{\bar{p}}^i$  contains the intrinsic step parameters and describes the reference configuration of the step (*i.e.*, the configuration of zero energy, also called the rest state), and  $\mathbf{F}^i$  is a  $6 \times 6$  matrix containing the elastic moduli associated with the different modes of deformation. Both the intrinsic step parameters and the force constant matrix can differ for each step depending on the DNA sequence. The elastic energy for the base-pair collection is then given by:

$$\mathcal{E} = \sum_{i=1}^{N-1} E^i. \quad (2)$$

The purpose of our method is to minimize the elastic energy of a base-pair collection given by Eq. (2) under the constraints detailed below. In other words, we need to calculate the derivatives of the elastic energy. The difficulty stems from the fact that this gradient has to be calculated with respect to a set of independent variables. This set of independent variables depends on the end conditions applied to the collection of base pairs and also on the presence of bound proteins. As mentioned above, the step parameters are not convenient for dealing with the end conditions. We therefore calculate the gradient with respect to the step dofs. We first consider the trivial case of an unconstrained collection (*i.e.*, a collection with no imposed end conditions and no bound proteins) and we then show how to account for specific end conditions and the presence of bound proteins.

Note that, our method does not require the elastic energy of a step to be a quadratic form with respect to the step parameters. It is also possible to include coupling between successive steps in the expression of the total elastic energy of the collection of base pairs.

### 2.3. Elastic energy gradient for a free base-pair collection

The case of a base-pair collection free of any end conditions and without bound proteins is trivial in the sense that the optimization leads to the reference configuration. We will use this gradient for a free collection as the starting point of our method.

The gradient of the elastic energy for a single step is obtained directly from Eq. (1):

$$\frac{\partial E^i}{\partial \underline{p}^i} = \mathbf{F}_s^i (\underline{p}^i - \bar{\underline{p}}^i), \quad (3)$$

where we introduce the symmetrized force constant matrix  $\mathbf{F}_s^i$  defined as  $\mathbf{F}_s^i = (\mathbf{F}^i + \mathbf{F}^{i\top})/2$ . It follows that the variation of the elastic energy for the complete collection is given by:

$$\delta \mathcal{E} = \frac{\partial \mathcal{E}}{\partial \underline{P}}^\top \delta \underline{P} = \sum_{i=1}^{N-1} \frac{\partial E^i}{\partial \underline{p}^i}^\top \delta \underline{p}^i = \sum_{i=1}^{N-1} (\mathbf{F}_s^i (\underline{p}^i - \bar{\underline{p}}^i))^\top \delta \underline{p}^i. \quad (4)$$

In order to obtain the gradient with respect to the step dofs we introduce the Jacobian matrix  $\mathbf{J}_\Phi$  defined as:

$$\delta \underline{P} = \frac{\partial \underline{P}}{\partial \Phi} \delta \Phi = \mathbf{J}_\Phi \delta \Phi. \quad (5)$$

It follows that:

$$\delta \mathcal{E} = \left( \mathbf{J}_\Phi^\top \frac{\partial \mathcal{E}}{\partial \underline{P}} \right)^\top \delta \Phi, \quad (6)$$

which leads to the following expression for the free-collection gradient with respect to the step dofs:

$$\frac{\partial \mathcal{E}}{\partial \Phi} = \mathbf{J}_\Phi^\top \frac{\partial \mathcal{E}}{\partial \underline{P}}. \quad (7)$$

The details of the calculation for the matrix  $\mathbf{J}_\Phi$  are given in Appendix B.2.

### 2.4. End conditions

We now consider the case of a base-pair collection with end conditions, that is, both the end-to-end vector and the orientation between the first and last base pairs are fixed. Our results can easily be generalized to other types of end conditions.

The end-to-end vector for the collection of base pairs is given by:

$$\underline{x}^N - \underline{x}^1 = \sum_{i=1}^{N-1} \underline{r}^i. \quad (8)$$

If the end-to-end vector is imposed, it follows that:

$$\sum_{i=1}^{N-1} \delta \underline{r}^i = \underline{0}, \quad (9)$$

and we can therefore write:

$$\delta \underline{r}^{N-1} = - \sum_{i=1}^{N-2} \delta \underline{r}^i. \quad (10)$$

The end-to-end rotation corresponds to the orientation between the first and last base pairs and is given by:

$$\mathbf{d}^1{}^\top \mathbf{d}^N = \mathcal{D}^{(1,N)} = \prod_{i=1}^{N-1} \mathcal{D}^i(\underline{\psi}^i), \quad (11)$$

where  $\mathcal{D}^{(i,j)}$  denotes the rotation matrix between the  $i$ -th and  $j$ -th base pairs, and  $\mathcal{D}^i$  is the rotation matrix for the  $i$ -th step which is completely parametrized by the angular step dofs  $\underline{\psi}^i = (\psi_1^i, \psi_2^i, \psi_3^i)$  (refer to Appendix A.1, Appendix B.1 and the reference [11] for more details). If the end-to-end rotation is imposed we have the constraint:

$$\delta\mathcal{D}^{(1,N)} = \delta\left(\prod_{i=1}^{N-1} \mathcal{D}^i(\underline{\psi}^i)\right) = \mathbf{0}. \quad (12)$$

This constraint can be written as a condition on  $\delta\underline{\psi}^{N-1}$ , that is:

$$\delta\underline{\psi}^{N-1} = -\sum_{i=1}^{N-2} \mathbf{K}^i \delta\underline{\psi}^i, \quad (13)$$

where the details of the matrix  $\mathbf{K}^i$  are given in Appendix C (see Eq. (C.7)).

It follows from the conditions given by Eq. (10) and Eq. (13) that the variation of the step dofs for the last step can be expressed as:

$$\delta\underline{\phi}^{N-1} = -\sum_{i=1}^{N-1} \mathbf{B}^i \delta\underline{\phi}^i, \quad (14)$$

where, once again, the details of the matrix  $\mathbf{B}^i$  are given in Appendix C (see Eq. (C.9)). This result means that the step dofs  $\underline{\phi}^{N-1}$  are not independent variables. In other words, the end conditions reduce the number of independent step dofs. Note that we choose to express the last step dofs as the non-independent variables but any other step could have been chosen. We now denote  $\hat{\underline{\Phi}}$  the set of independent step dofs and we have the relation  $\underline{\Phi} \subset \hat{\underline{\Phi}}$ . We then define the Jacobian matrix  $\hat{\mathbf{J}}$  such that:

$$\delta\underline{\Phi} = \hat{\mathbf{J}} \delta\hat{\underline{\Phi}}. \quad (15)$$

The dimensions of this matrix depend on the precise details of the end conditions (although the number of rows is always  $6(N-1)$ ) and the details of its expression are given in Appendix C.

The variation of the elastic energy for a collection of base pairs subjected to imposed end conditions can be written as:

$$\delta\mathcal{E} = \left( \left( \mathbf{J}_{\underline{\Phi}} \hat{\mathbf{J}} \right)^\top \frac{\partial \mathcal{E}}{\partial \underline{P}} \right)^\top \delta\hat{\underline{\Phi}}. \quad (16)$$

It follows that the gradient is:

$$\frac{\partial \mathcal{E}}{\partial \hat{\underline{\Phi}}} = \left( \mathbf{J}_{\underline{\Phi}} \hat{\mathbf{J}} \right)^\top \frac{\partial \mathcal{E}}{\partial \underline{P}}. \quad (17)$$

Note that this gradient corresponds to the gradient with respect to the independent step dofs and accounts for the end conditions. This is one advantage of our method: the end conditions are directly accounted for. Hence, we transform a constrained optimization problem into an unconstrained one, which makes the numerical implementation simpler and more robust.

### 2.5. Bound-protein constraint

In order to study protein-decorated DNA we need to account for the presence of bound proteins on the double helix. We model the binding of proteins on the DNA by considering the step parameters of the binding domain as *frozen*, that is, these step parameters are imposed and cannot change. For example, the frozen step parameters can be extracted from high-resolution crystal structures of protein-DNA complexes with the help of the 3DNA software [16].

We now consider that the  $k$ -th step in the base-pair collection is frozen, that is, the step parameters  $\underline{p}^k$  are imposed (the following results can be easily generalized to an arbitrary number of frozen steps). It follows directly that the angular step dofs are constant and we therefore have  $\delta\underline{\psi}^k = \underline{0}$ . The translational step dofs, however, are not constant: any changes in the angular steps dofs  $\underline{\psi}^j$  with  $j < k$  will change the

orientation of the joining vector  $\underline{r}^k$  of the frozen step. Indeed, we show in Appendix D that the variations of the translational step dofs of the frozen step can be written as:

$$\delta \underline{r}^k = \sum_{j=1}^{k-1} \mathbf{W}^{j,k} \delta \underline{\psi}^j. \quad (18)$$

We therefore obtain for the variation of the frozen step dofs:

$$\delta \underline{\phi}^k = \sum_{j=1}^{k-1} \mathbf{C}^{j,k} \delta \underline{\phi}^j, \quad (19)$$

where the details of the matrices  $\mathbf{W}^{j,k}$  and  $\mathbf{C}^{j,k}$  are given in Appendix D (see Eq. (D.4)).

Similar to the set of independent step dofs  $\hat{\underline{\Phi}}$  introduced for the treatment of the end conditions, we introduce a new set of independent step dofs  $\tilde{\underline{\Phi}}$  which corresponds to the set of non-frozen step dofs among the set  $\hat{\underline{\Phi}}$ , that is, we have  $\tilde{\underline{\Phi}} \subset \hat{\underline{\Phi}} \subset \underline{\Phi}$ . We define the matrix  $\tilde{\mathbf{J}}$  as the Jacobian (see Appendix D for details):

$$\delta \hat{\underline{\Phi}} = \tilde{\mathbf{J}} \delta \tilde{\underline{\Phi}}. \quad (20)$$

We obtain for the variation of the energy:

$$\delta \mathcal{E} = \left( \left( \tilde{\mathbf{J}}_{\underline{\Phi}} \hat{\mathbf{J}} \tilde{\mathbf{J}} \right)^{\top} \frac{\partial \mathcal{E}}{\partial \tilde{\underline{P}}} \right)^{\top} \delta \tilde{\underline{\Phi}}, \quad (21)$$

and we have for the gradient:

$$\frac{\partial \mathcal{E}}{\partial \tilde{\underline{\Phi}}} = \left( \tilde{\mathbf{J}}_{\underline{\Phi}} \hat{\mathbf{J}} \tilde{\mathbf{J}} \right)^{\top} \frac{\partial \mathcal{E}}{\partial \tilde{\underline{P}}}. \quad (22)$$

In the above expression,  $\tilde{\underline{P}}$  denotes the set of non-frozen step parameters and  $\tilde{\mathbf{J}}_{\underline{\Phi}}$  is the Jacobian matrix defined as:

$$\tilde{\mathbf{J}}_{\underline{\Phi}} = \frac{\partial \tilde{\underline{P}}}{\partial \tilde{\underline{\Phi}}}. \quad (23)$$

This matrix is directly obtained from  $\mathbf{J}_{\underline{\Phi}}$  by removing the rows associated with the frozen step parameters.

The gradient obtained with Eq. (22) corresponds to the gradient of the elastic energy of the base-pair collection (Eq. (2)) with respect to the set of non-frozen independent step dofs.

## 2.6. Force field

We introduced in the expression of the elastic energy of a base-pair step (Eq. (1)) two quantities which depend on the sequence of the DNA base-pair collection: the intrinsic step parameters  $\underline{p}^i$  and the force constants matrix  $\mathbf{F}^i$ . These quantities constitute the force field of our theory. The intrinsic step parameters describe the rest states of the base-pair steps, while the force constants are the stiffnesses of the steps.

In this work, we use a uniform ideal force field (*i.e.*, not depending on the details of the DNA sequence). The intrinsic step parameters for this force field are given by  $\underline{p} = (0 \text{ deg}, 0 \text{ deg}, 34.2857 \text{ deg}, 0 \text{ \AA}, 0 \text{ \AA}, 3.4 \text{ \AA})$ , which correspond to a B-DNA like straight rest state and a helical repeat of 10.5 bp. We do not consider coupling between the different modes of deformation and, hence, the force constants matrix is diagonal and reads:

$$\text{diag}(\mathbf{F}^{\text{ideal}}) = (0.0427, 0.0427, 0.0597, 20, 20, 20). \quad (24)$$

The units for the the first three diagonal entries are  $k_B T \cdot \text{deg}^{-2}$  and  $k_B T \cdot \text{\AA}^{-2}$  for the last three entries. This force field can be considered as quasi-inextensible due to the high force constants associated with the translational step parameters. In the limit of a infinitely long base-pair collection we can calculate the persistence lengths corresponding to the force field and we find that the bending and twisting persistence lengths are 47.7 nm and 66.6 nm, respectively.

### 2.7. Protein binding procedure

As explained earlier, we model the presence of proteins on DNA by setting the base-pair step parameters of the binding domain to imposed values. These imposed values can be extracted from high-resolution crystal structures of protein-DNA complexes or can be chosen to represent an ideal protein-DNA binding system.

In this study, we use an iterative procedure to bind a protein onto DNA progressively. For example, we consider a protein with a binding domain of  $n_p$  steps to be bound starting at the  $k$ -th step, that is, the steps  $k$  to  $k + n_p - 1$  are bound (and, hence, frozen). The iterative procedure consists in adjusting the step parameters of the binding domain using the following *linear ramp* function:

$$\underline{p}^j(\lambda) = (1 - \lambda)\underline{p}_{\text{free}}^j + \lambda\underline{p}_{\text{protein}}^j, \forall j \in [k, k + n_p - 1], \quad (25)$$

where  $\underline{p}_{\text{free}}^j$  denote the step parameters of the protein-free DNA (in our case they will correspond to the step parameters of a portion of a naked-DNA minicircle) and  $\underline{p}_{\text{protein}}^j$  are the step parameters of the DNA found in the protein-DNA complex. The parameter  $\lambda \in [0, 1]$  is the ramp parameter. Our procedure starts with  $\lambda = 0$  and we minimize the DNA base-pair collection while gradually increasing the value of the ramp parameter until  $\lambda = 1$ . As an outcome we obtain an optimized DNA base-pair collection in which a part of DNA is *shaped* as if the protein were bound to it. This numerical procedure is not meant to convey the physical process of proteins binding to DNA. It is designed to set regions of DNA to the states found in protein-DNA complexes. We also would like to point out that the binding ramp can be tuned in order to improve the robustness of the method. For example, the linear terms can be replaced by quadratic expressions. Such modifications might be needed to avoid instabilities or to control changes in the total twist while ramping the step parameters in the case of binding domains with large deformations.

### 2.8. Minicircle topology

The topology of a DNA minicircle is characterized by its linking number Lk [17], writhing number Wr [18, 19], and total twist Tw [18]. The linking number of a minicircle, an integer, describes the entanglement of the minicircle centerline and the curve traced out by one of the edges of the base-pair frames (see [20] for detailed explanations and computational methods). In particular, for a planar minicircle the linking number corresponds to the number of turns the double helix makes. For a minicircle of  $N$  bp, the *relaxed* linking number  $\text{Lk}^0$  is given by the integer nearest to  $N/10.5$ , where 10.5 is the assumed helical repeat of DNA. We introduce the difference between the actual and relaxed linking numbers of a minicircle as  $\Delta\text{Lk} = \text{Lk} - \text{Lk}^0$ . Because a DNA minicircle is covalently closed, its linking number, and hence  $\Delta\text{Lk}$ , are constant and are not altered by the deformation of the double helix.

The linking number of a minicircle is always equal to the sum of its writhing number and total twist,  $\text{Lk} = \text{Tw} + \text{Wr}$ . The writhing number characterizes the global folding and non-chiral distortions of the minicircle centerline, while the total twist measures the twisting or *twist density* of the base pairs around the centerline. Like the linking number, the writhing number and total twist can be directly obtained from the minicircle base-pair frames as explained in [20]. The invariance of the linking number implies that when the minicircle is deformed, the total twist and the writhing number are redistributed. The total twist is directly related to the torsional stress within the double helix, while the writhing number depends on the curvature of the centerline, albeit in a non-trivial way.

### 2.9. Protein-free DNA minicircles

In order to study protein-decorated DNA minicircles we first need to generate protein-free minicircles. For a minicircle of  $N$  bp, we first build a set of step parameters using the following formula for the  $i$ -th step:

$$\underline{p}^i = (\Delta \sin(\theta_3^* i), \Delta \cos(\theta_3^* i), \theta_3^*, 0, 0, \bar{\rho}_3), \quad (26)$$

where  $\Delta = 2\pi/N$  is the bending angle between the planes of successive base pairs and  $\bar{\rho}_3$  denotes the intrinsic value of the rise step parameter within our force field. The parameter  $\theta_3^*$  corresponds to the local

twist density in the minicircle and can be adjusted to generate over- or undertwisted configurations. In order to enforce the covalent closure of the minicircle, we add at the end of the base-pair collection an additional base-pair identical to the first one. We then minimize the energy of the base-pair collection under the imposed end-to-end vector and rotation (which are both null in the present case). The outcome of the minimization is an optimized DNA minicircle with an imposed  $\Delta Lk$ .

Here we initially focus on planar, protein-free minicircles of lengths ranging from 63 bp to 105 bp (*i.e.*, from 6 to 10 helical repeats). Planar minicircles always have a writhing number equal to zero, which means that the total twist is equal to the linking number. Because we require the naked minicircles to be planar there is an upper bound on the value of  $|Tw|$  (and, hence, of  $|\Delta Lk|$ ). This limiting value is related to the Michell-Zajac instability [21, 22], and for larger values of  $|Tw|$  the planar circular configurations are no longer stable. With our ideal force field, for a given chain length we can always generate optimized minicircles with  $\Delta Lk = 0$  and, depending on the chain length,  $\Delta Lk = -1$  or  $\Delta Lk = +1$  (for a few specific chain lengths the three values are possible; see Appendix E for more details). In other words, we always compute optimized protein-free minicircles with  $\Delta Lk = 0$  and in addition, we also generate optimized minicircles with  $\Delta Lk = -1$  and/or  $\Delta Lk = +1$ . These different minicircles are then used with the binding ramp (Eq. (25)) to model the presence of bound proteins.

### 3. Results

We first apply our minimization method to the optimization of DNA minicircles on which one or two proteins are bound. We focus on two distinct proteins: the abundant histone-like HU protein and the structurally related Hbb protein. These two proteins have a common fold and share some structural similarities, such as the fact that they both associate as dimers and both introduce significant localized bends in the double helix. We used the 3DNA software [16] to extract the step parameters of the DNA found in known crystal complexes [1, 5] (Protein Data Bank files 1P71 and 2NP2 for HU and Hbb, respectively). The lengths of the extracted binding domains are 17 bp for HU and 35 bp for Hbb. Within the scope of our work, we think of the Hbb-bound DNA as an extreme model of HU-induced distortion.

The net bending angle introduced by the HU dimer in the selected structure is roughly 135 deg, and that by Hbb is 160 deg (these angles correspond to the angles between the normals of the first and last base pairs). Both proteins also slightly underwind the double helix. That is, the total twist across the DNA bound to each protein is less than the total twist of an undeformed DNA fragment of same length (for HU  $\Delta Tw \simeq -0.18$  turns and for Hbb  $\Delta Tw \simeq -0.1$  turns).

#### 3.1. Minicircles with a single bound protein

We first performed a series of optimizations to study how the addition of an HU or Hbb dimer affects a DNA minicircle. We focused on chain lengths of 63 bp to 105 bp and considered relaxed minicircles ( $\Delta Lk = 0$ ) as well as over- or undertwisted minicircles ( $\Delta Lk = \pm 1$ ).

In order to characterize these optimized configurations we define a relative step energy  $\epsilon$  as:

$$\epsilon = \frac{N}{N_{\text{free}}} \frac{\mathcal{E}}{\mathcal{E}^0}, \quad (27)$$

where  $\mathcal{E}$  is the energy of the optimized minicircle with the bound dimer,  $\mathcal{E}^0$  is the energy of the optimized protein-free minicircle,  $N$  denotes the minicircle chain length, and  $N_{\text{free}}$  the protein-free chain length. In other words,  $\epsilon$  is the ratio of the energy per base-pair step for the minicircle with a single protein versus that for the naked minicircle. We recall that our elastic energy for a minicircle accounts for protein-free steps only. The results for HU are presented in Fig. 1 and those for Hbb in Fig. 2.

We first remark that the presence of an HU dimer almost always reduces the energy of the protein-free DNA compared to that of the naked minicircle (as shown by the values of the relative step energy  $\epsilon$  less than one in Fig. 1). On the other hand, the addition of an Hbb protein has a mixed effect on the relative step energy depending on the chain length and the value of  $\Delta Lk$  (Fig. 2). A possible explanation lies in the fact that the binding domain of the Hbb dimer is twice as long as that for the HU dimer. Thus, the



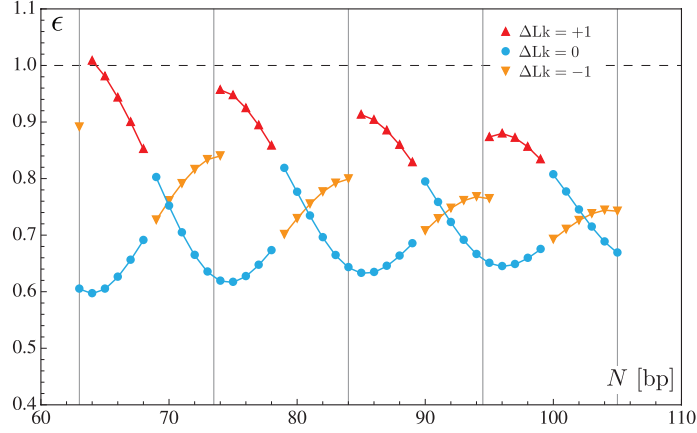


Figure 1: Chain-length dependence of the relative step energy  $\epsilon$  for minicircles with a single HU dimer. The blue circles correspond to the energies of relaxed minicircles and span the whole range of chain lengths. The red and orange triangles represent the energies of over- and underwound minicircles, respectively. The horizontal dashed line ( $\epsilon = 1$ ) corresponds to the energy of a base-pair step in a naked minicircle of the same length. The vertical lines indicate the chain lengths equal to integral numbers of helical repeats.

length of protein-free DNA is less for minicircles with an Hbb protein, although the boundary conditions are comparable to those of HU. In other words, the deformation required to satisfy the boundary conditions is larger in minicircles with an Hbb dimer, and, hence, the energy is higher. This argument also accounts for the differences in the chain-length dependence of the relative step energy for minicircles containing an HU dimer versus those with an Hbb dimer. As shown in Fig. 1, for the HU dimer the chain-length dependence of the relative step energy follows a damped oscillatory pattern (the period is roughly equal to the assumed 10.5-bp DNA helical repeat). By contrast, there is no clear periodicity in the Hbb results (Fig. 2). This suggests, that within the present range of chain lengths, a minicircle with a single Hbb dimer is more *constrained* than one with an HU protein.

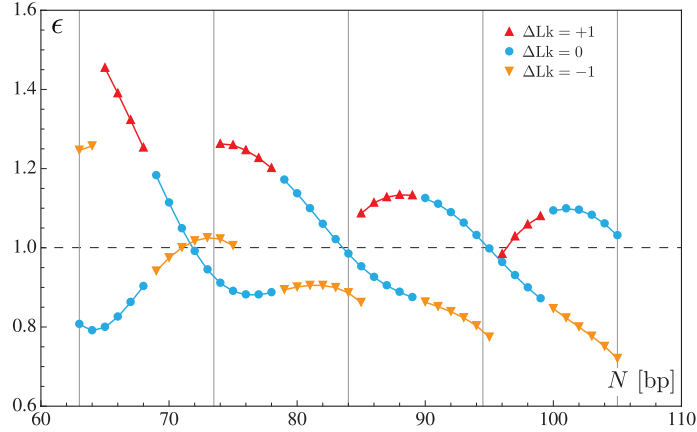


Figure 2: Chain-length dependence of the relative step energy  $\epsilon$  for minicircles with a single Hbb dimer. The blue circles correspond to the energies of relaxed minicircles and span the whole range of chain lengths. The red and orange triangles represent the energies of over- and underwound minicircles, respectively. The horizontal dashed line ( $\epsilon = 1$ ) corresponds to the energy of a base-pair step in a naked minicircle of the same length. The vertical lines indicate the chain lengths equal to integral numbers of helical repeats.

The other interesting feature found in these results is the influence of the linking number and the total twist on the relative step energy. In the case of the HU dimer, relaxed minicircles have, in most cases, a lower energy than under- or overwound minicircles. It is only for chain lengths close to half-

integral numbers of helical repeats that the energy is lower for underwound minicircles. On the other hand, underwound minicircles of 79 bp or greater bearing an Hbb dimer are consistently lower in energy than relaxed or overwound minicircles. As mentioned earlier, both dimers slightly underwind the double helix, which means that their presence on a minicircle induces a redistribution of the torsional stress. Note that, this redistribution may entail the conversion of part of the torsional stress into bending deformations as a consequence of the changes in the total twist and the writhing number. This also explains why overwound minicircles lead to higher energies: the large difference in the torsional stress between the naked DNA (prior to the addition of a protein) and the bound DNA causes higher deformations in the remaining part of the minicircle. In addition, the torsional stress in naked DNA depends on the chain length: for chain lengths close to integral numbers of helical repeats, the torsional stress is lower in relaxed minicircles than in underwound minicircles, but for chain lengths near half-integral numbers of helical repeats the situation is reversed.

Our results show that the torsional stress in DNA may influence the recruitment of HU and Hbb dimers and might act as a *control mechanism* for the presence of such proteins. For example, a relaxed minicircle of 94 bp is more likely to be bound to an HU than an Hbb dimer (assuming that the binding affinity of both dimers are comparable), whereas an underwound minicircle of the same length is more likely to take up an Hbb dimer. Although our results are obtained on covalently closed DNA minicircles (and, hence, under a topological constraint), it is reasonable to think that similar effects will take place on torsionally constrained linear DNA fragments (for example, the anchoring conditions could be achieved by large protein assemblies or magnetic/optical tweezers).

### 3.2. Minicircles with two bound proteins

Our second series of optimizations focuses on minicircles of 100 bp and 105 bp containing two HU or two Hbb dimers. In addition, for minicircles of 100 bp we consider relaxed ( $\Delta Lk = 0$ ) and underwound ( $\Delta Lk = -1$ ) molecules. This choice is motivated by the fact that the difference in energy between a naked DNA minicircle of 100 bp with  $\Delta Lk = 0$  and  $\Delta Lk = -1$  is only of 1.8 k<sub>B</sub>T (the relaxed minicircle corresponds to the lower energy). Our computations provide energy landscapes of the protein-bound minicircles as functions of the spacing between the two binding sites. These landscapes are directly related to the relative likelihoods of forming minicircles with two dimers at specific locations. In other words, the landscapes reveal which dimer positions along the minicircle are more apt to be occupied in, for example, cyclization experiments. We report in Figs. 3-6 the energy landscapes as well as the associated changes in the total twist  $\Delta Tw$  (with respect to the planar and naked minicircles).

The energy landscapes consist of several local minima separated by high energy states. In other words, there are well-defined locations for optimal placements of two HU or Hbb proteins along the DNA minicircles. Notice that, only the minima with the lowest energies and, hence, with the highest Boltzmann weights, are likely to be relevant to the statistical physics of protein-decorated minicircles. These minima appear periodically in the landscapes and, as expected, the period is roughly equal to the assumed DNA helical repeat (10.5 bp). In addition, the relative locations of the two proteins along the minicircle appear to affect the torsional stress significantly as evidenced by the variations in the total twist. We also notice that the lowest energy configurations are all similar from a geometric point of view. That is, for both HU and Hbb proteins the globally optimal configurations are obtained when the two dimers are located at antipodal or near antipodal sites (as shown by the configurations labeled ‘1’ in Fig. 3 to Fig. 6). We note that the minima for the minicircles with two HU dimers are of lower energy than those for minicircles with two Hbb dimers. For Hbb-bound minicircles the length of protein-free DNA is shorter than for HU (for two Hbb dimers, the length of naked DNA is comparable to the size of the Hbb binding domain). The Hbb system is therefore highly constrained and leads to larger values in the energy landscapes and greater changes in the total twist.

The results for the minicircles of 100 bp containing two HU or two Hbb dimers (see Figs. 3 and 5) show that, the energy minima are lower for the underwound minicircle ( $\Delta Lk = -1$ ) than for the relaxed minicircle ( $\Delta Lk = 0$ ). This is consistent with the results obtained for a minicircle of 100 bp with a single dimer (see Figs. 1 and 2). We remark that the variations in the energy landscapes of the relaxed and underwound minicircles of 100 bp are out of phase. That is, the minima for the relaxed minicircles correspond to the maxima of the underwound minicircles and *vice versa*. This *phase shift* between the energy landscapes of

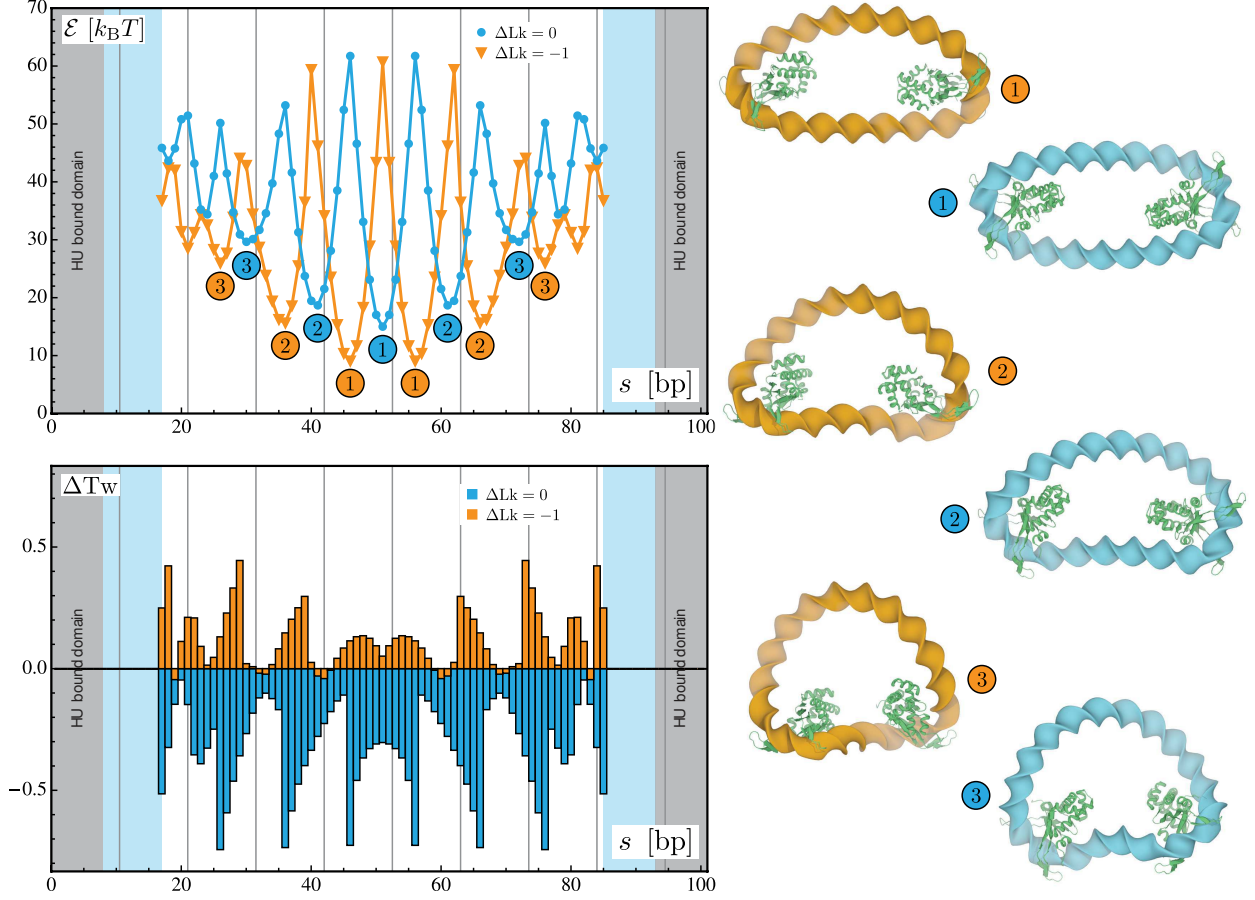


Figure 3: Optimization results for relaxed and underwound minicircles of 100 bp with two HU dimers. The two plots on the left represent (top) the optimized energy and (bottom) the changes in the total twist as functions of the center-to-center spacing  $s$  between the two proteins. In both plots, the gray areas denote the binding domain of the first HU protein and the light blue areas the binding domain of the second protein. The vertical lines indicate the chain lengths equal to integral numbers of helical repeats. The numbers in the energy plot refer to the structures depicted on the right in which the HU proteins are represented in green. The underwound ( $\Delta Lk = -1$ ) and relaxed ( $\Delta Lk = 0$ ) minicircles are represented in orange and blue, respectively.

relaxed and underwound minicircles is roughly equal to half the assumed helical repeat ( $\sim 5-6$  bp). The changes in the total twist for these minicircles are also similar, although the magnitude of the change is larger for the minicircles with two Hbb proteins. For the relaxed minicircles of 100 bp the changes in the total twist are negative (with respect to the planar, protein-free minicircles), while for underwound minicircles the changes are positive. This implies, because of the conservation of the linking number, that the changes in the writhing number are of opposite sign for relaxed versus underwound minicircles. Indeed, as shown in Fig. 5 (configurations labeled ‘2’), the minicircles are of different *handedness*. We also notice that, the pattern in the changes in the total twist for minicircles of 100 bp with two dimers is similar: the local maxima in the energy correspond to the highest changes in the total twist, while the local minima correlate with those of moderate values of  $\Delta Tw$ .

For the minicircles of 105 bp containing two HU or Hbb dimers, the energy landscapes are similar to those obtained for 100-bp minicircles. In particular, the minima appear periodically (every  $\sim 10.5$  bp) and are separated by high energy configurations. The minima for the 105-bp chains are lower than those on the minicircles of 100 bp. This difference is due to the fact that in 105-bp minicircles the torsional stress in the minicircle prior to the addition of the proteins is lower than that in the 100-bp minicircles. The main

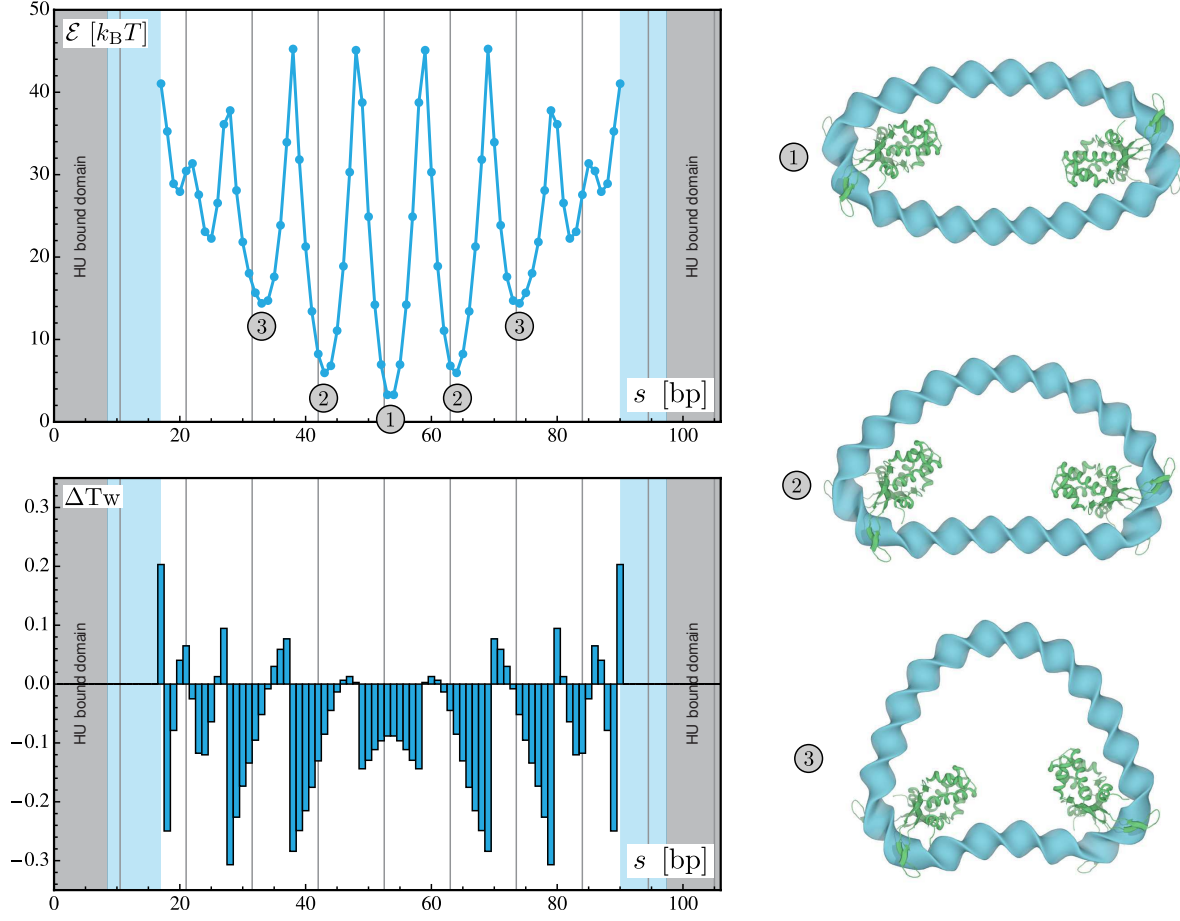


Figure 4: Optimization results for minicircles of 105 bp with two HU dimers. The two plots on the left represent (top) the optimized energy and (bottom) the changes in the total twist as functions of the center-to-center spacing  $s$  between the two proteins. In both plots, the gray areas denote the binding domain of the first HU protein and the light blue areas the binding domain of the second protein. The vertical lines indicate the chain lengths equal to integral numbers of helical repeats. The numbers in the energy plot refer to the structures depicted on the right in which the HU proteins are represented in green.

dissimilarity between minicircles of 100 bp and 105 bp resides in the changes in the total twist. We have seen that for 100-bp minicircles, the changes in the total twist are comparable for HU and Hbb dimers. In the case of the minicircles of 105 bp, however, the values of  $\Delta Tw$  corresponding to the local minima are of different signs for the Hbb and HU minicircles. That is, for the lower energy configurations, the presence of two HU dimers reduces the torsional stress in the minicircle (see Fig. 4), while two Hbb dimers increase the torsional stress (see Fig. 6). We also notice that, for these local minima the magnitude of the changes in the total twist is comparable.

Our study of minicircles with two dimers reveals the interplay between the elastic properties of DNA and the positioning of proteins on the double-helix. It appears that this is not about the proteins shaping the double helix nor the DNA stiffnesses controlling the positioning of proteins. Rather there is cooperation between the presence of proteins and the elasticity of the double helix, particularly in the distribution of the torsional stress. The high degree of contrast in our energy landscapes suggests that, once a protein is bound to a topologically constrained DNA fragment, the stress in the double helix will favor specific binding sites for other proteins. Such an interpretation echoes the recent results about DNA-protein allosteric effects obtained in single-molecule experiments [23]. It is also interesting to notice that the local minima found in our energy landscapes are always flanked by configurations of comparable energy (up to a few  $k_B T$ ). This

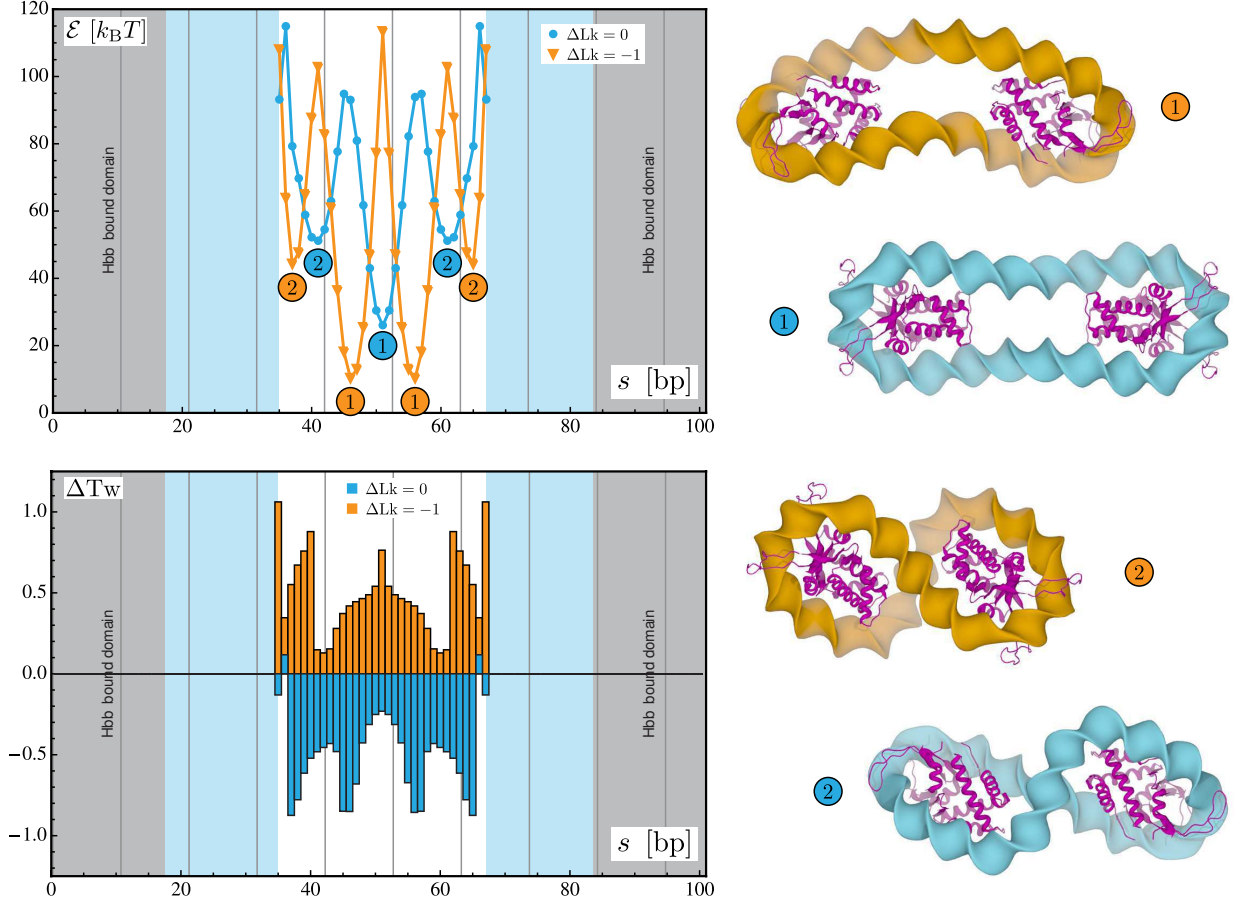


Figure 5: Optimization results for relaxed and underwound minicircles of 100 bp with two Hbb dimers. The two plots on the left represent (top) the optimized energy and (bottom) the changes in the total twist as functions of the center-to-center spacing  $s$  between the two proteins. In both plots, the gray areas denote the binding domain of the first Hbb protein and the light blue areas the binding domain of the second protein. The vertical lines indicate the chain lengths equal to integral numbers of helical repeats. The numbers in the energy plot refer to the structures depicted on the right in which the Hbb proteins are represented in pink. The underwound ( $\Delta Lk = -1$ ) and relaxed ( $\Delta Lk = 0$ ) minicircles are represented in orange and blue, respectively.

suggests that there should be fluctuations in the experimental measurements of the most likely binding sites of HU and Hbb dimers along DNA minicircles. Notice that, our results have been obtained with a minimal model, that is, DNA is modeled as an isotropic material with standard bending and twisting stiffnesses and we have focused on a special class of protein geometry which includes specific deformations on the double helix. Nevertheless, our approach serves as a proof of concept and paves the way for more detailed studies about the synergy between DNA deformation and protein positioning.

#### 4. Discussion

The minimization procedure introduced in this work facilitates the investigation of how architectural proteins may contribute to the spatial organization and genetic processing of DNA. This new approach gives us direct control of the positions and orientations of the base pairs at the ends of a DNA chain and allows us to specify the precise sites of protein uptake and the detailed changes in double-helical structure brought about by the binding of protein. Here we illustrate the utility of the method in a study of the elastic energy landscapes of DNA minicircles decorated by the nonspecific architectural protein HU and

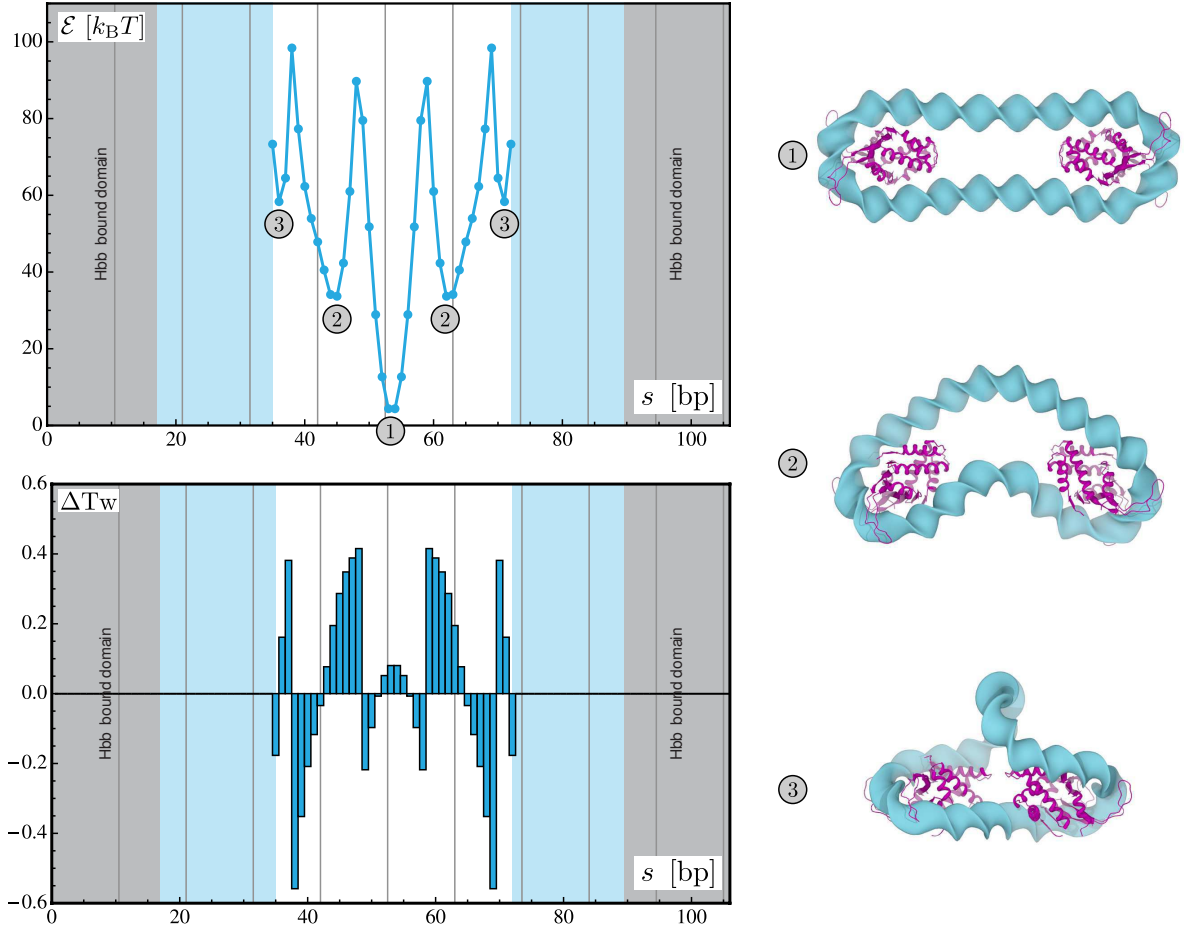


Figure 6: Optimization results for minicircles of 105 bp with two Hbb dimers. The two plots on the left represent (top) the optimized energy and (bottom) the changes in the total twist as functions of the center-to-center spacing  $s$  between the two proteins. In both plots, the gray areas denote the binding domain of the first Hbb protein and the light blue areas the binding domain of the second protein. The vertical lines indicate the chain lengths equal to integral numbers of helical repeats. The numbers in the energy plot refer to the structures depicted on the right in which the Hbb proteins are represented in pink.

the similarly folded, albeit site-specific [24], Hbb protein. Both proteins associate as dimers and introduce severe bends and untwist their DNA targets. We consider the Hbb-bound DNA as an extreme example of HU-induced DNA distortion and thus treat both proteins as nonspecific. We focus on covalently closed molecules comparable in length to the loops that are formed by various regulatory proteins and enzymes that bind to sequentially distant sites on DNA [25, 26] and allow for the uptake of one or two HU or Hbb dimers on the DNA. We also consider underwound and overwound minicircles in order to study the added effects of the torsional stress within the double helix on the protein-binding landscapes. We find that the presence of protein has a significant effect on the bending and twisting deformations in the minicircles, and conversely, that the torsional stress within DNA prior to the addition of proteins has a strong effect on the optimal placement of proteins along the minicircles. For example, we show that a single HU dimer is more likely to bind relaxed rather than under- or overwound minicircles of most chain lengths between 63-105 bp and that an Hbb dimer binds preferentially to underwound minicircles of the same lengths.

Our results reveal cooperation between the deformability of the double helix and the structural distortions of DNA induced by bound proteins. In the case of minicircles with two HU or two Hbb dimers, the presence of a first protein strongly influences the locations of the optimal binding sites of a second protein. That is, the DNA, through its elastic deformation, acts as a communication medium between the proteins. In



particular, the torsional stress and the twisting stiffness appear to play a major role in this action at a distance. The mechanical signaling also provides a rationale for the DNA allostery reported in recent single-molecules studies of the dissociation of proteins on DNA chains constrained to full extension by the flow of solvent [23, 27]. In particular, the binding of one protein on the extended DNA stabilizes or destabilizes the binding of another protein, even when not in direct contact. Indeed, this sort of long-range communication, in which the binding of a ligand in one part of the DNA helix influences (positively or negatively) the recognition of a different ligand at a remote site, also termed telestability [28], has puzzled DNA scientists for decades. Although our results concern a special class of proteins bound to covalently closed rather than straightened DNA, we observe an interplay between the elasticity of the double helix and the placement of proteins along DNA that may hold for extended as well as cyclic chains. In our case, the most likely placement of a second protein is antipodal to the first, *i.e.*, as sequentially far apart as possible. Our studies provide examples of how the mechanical stress in DNA can control the placement of proteins and how proteins can alter the mechanical stress to broadcast their presence along DNA. Our findings also suggest that local modifications of the mechanical properties of DNA, such as methylation or the occurrence of kinks, could modulate and possibly repress this type of mechanical signaling.

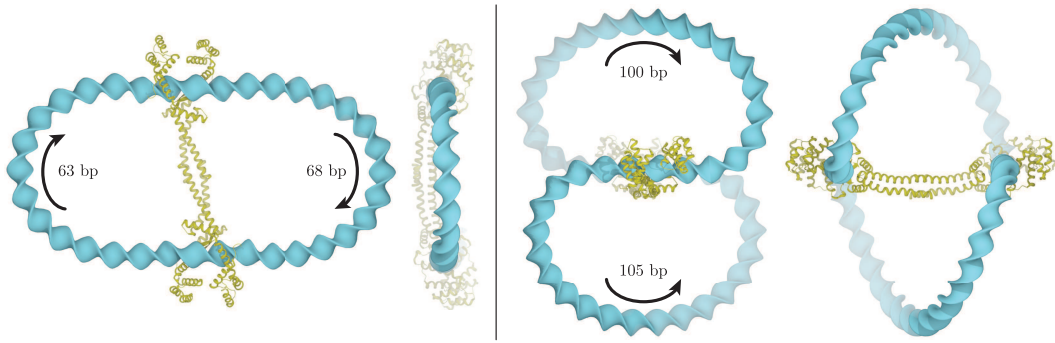


Figure 7: Orthogonal views of optimized loops induced by the DNA-bridging protein MatP [7] on circular DNA. The loops depicted on the left side run in anti-parallel directions and those on the right side run in parallel directions (note the arrows and chain lengths in the images). The total chain lengths of the circular DNA on the left and right sides are 177 bp and 251 bp, respectively. The MatP dimeric protein is represented in yellow. The step parameters of the two bound DNA domains have been extracted from Protein Data Bank file 3VEA.

Although the minimal elastic energy configurations of long DNA chains are not necessarily relevant from a statistical physics point of view, our method can be used to study larger systems, such as long DNA molecules decorated by multiple proteins. For example, we show in Fig. 7 two examples of multiple loops induced by the Ter-specific protein MatP [7] on circular DNA, which is thought to be involved in the condensation of chromosomes in *Escherichia coli*. We also show in Fig. 8 a long DNA plasmid decorated by 64 closely spaced Hbb proteins. Our software can also be used together with other tools (for example, 3DNA [16]) to optimize DNA fragments anchored by proteins. An interesting application of our approach resides in the development of software for biomolecular sculpting [29], which makes it possible to study how proteins can bundle and organize DNA. We are currently working on improving our method to account for additional types of constraints, such as the treatment of excluded volume. To date, our software does not check for collisions between DNA and proteins. Although the proteins collide on some of the minicircles presented in this work (see Figs. 3-6), the collisions only occur in high-energy configurations, in which pairs of proteins lie immediately next to one another. We also plan to include more realistic force fields to account, for example, for the higher deformability of pyrimidine-purine compared to other base-pair steps and for the flexibility of protein assemblies. This task is facilitated by the fact that our method already accounts for the sequence-dependent elasticity of DNA. Finally, we are now concentrating our efforts on other biomolecular systems, including the loops mediated by the Lac and Gal repressor proteins and the relative contributions of protein and DNA flexibility in loop formation.

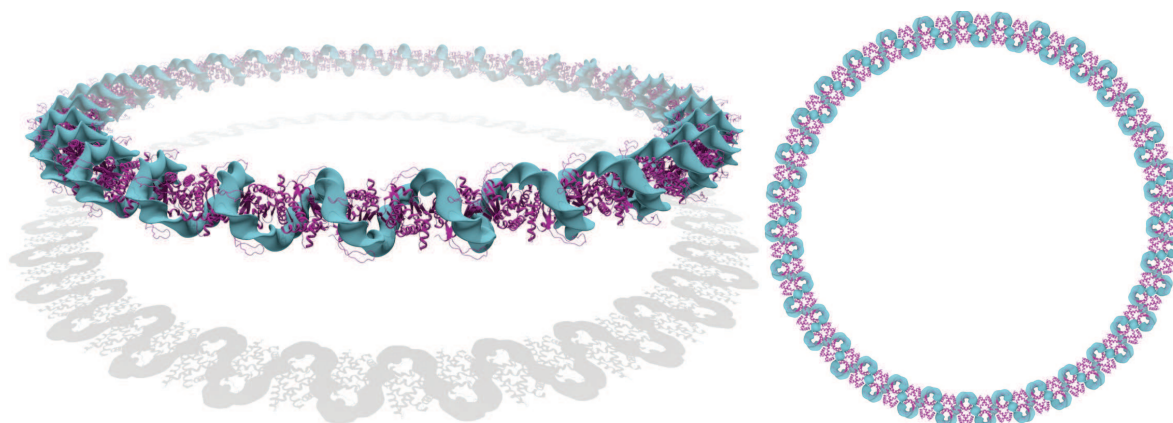


Figure 8: Perspective (left) and orthographic (right) views of an optimized DNA plasmid of 1280 bp containing 64 Hbb proteins regularly spaced by 5-bp linkers. The presence of the Hbb proteins (represented in pink) greatly condenses the DNA into a helical *fiber-like* structure that folds into a circular configuration. The circular radius of the resulting superstructure is 31.5 nm and the fiber radius is 3 nm (the same DNA plasmid without any proteins has a circular radius of 69.2 nm and a cross-sectional radius of 1 nm). The elastic energy stored in each naked-DNA linkers is 0.19  $k_B T$ , which suggests that the linkers are lowly deformed. In contrast to the zigzag superhelical structure found in the crystal lattice (see Fig. 1c in [5]) in which the proteins lie on opposite sides of the superhelical axis, the proteins form a central core in the optimized structure. Notice that, this structure has been obtained with a *reduced* Hbb model in which only the fifteen central base pairs of the full model described earlier are used. As a consequence, the spacing between the proteins is greater in the optimized structure than in the pseudo-continuous helix found in the crystal lattice.

## References

- [1] K. K. Swinger, K. M. Lemberg, Y. Zhang, P. A. Rice, Flexible DNA bending in HU-DNA cocrystal structures, *The EMBO Journal* 22 (14) (2003) 3749–3760.
- [2] K. K. Swinger, P. A. Rice, IHF and hu: flexible architects of bent DNA, *Current Opinion in Structural Biology* 14 (1) (2004) 28–35.
- [3] K. K. Swinger, P. A. Rice, Structure-based Analysis of HU-DNA Binding, *Journal of Molecular Biology* 365 (4) (2007) 1005–1016.
- [4] D. Sagi, N. Friedman, C. Vorgias, A. B. Oppenheim, J. Stavans, Modulation of DNA Conformations Through the Formation of Alternative High-order HU-DNA Complexes, *Journal of Molecular Biology* 341 (2) (2004) 419–428.
- [5] K. W. Mouw, P. A. Rice, Shaping the *Borrelia burgdorferi* genome: crystal structure and binding properties of the DNA-bending protein Hbb, *Molecular Microbiology* 63 (5) (2007) 1319–1330.
- [6] S. T. Arold, P. G. Leonard, G. N. Parkinson, J. E. Ladbury, H-NS forms a superhelical protein scaffold for dna condensation, *Proceedings of the National Academy of Sciences* 107 (36) (2010) 15728–15732.
- [7] P. Dupaigne, N. K. Tonthat, O. Espéli, T. Whitfill, F. Boccard, M. A. Schumacher, Molecular Basis for a Protein-Mediated DNA-Bridging Mechanism that Functions in Condensation of the *E. coli* Chromosome, *Molecular Cell* 48 (4) (2012) 560–571.
- [8] R. T. Dame, O. J. Kalmykova, D. C. Grainger, Chromosomal macrodomains and associated proteins: Implications for DNA organization and replication in gram negative bacteria, *PLoS Genetics* 7 (6) (2011) e1002123.
- [9] K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent, T. J. Richmond, Crystal structure of the nucleosome core particle at 2.8-Å resolution, *Nature* 389 (6648) (1997) 251–260.
- [10] A. G. West, M. Gaszner, G. Felsenfeld, Insulators: many functions, many mechanisms, *Genes & Development* 16 (3) (2002) 271–288.
- [11] B. D. Coleman, W. K. Olson, D. Swigon, Theory of sequence-dependent DNA elasticity., *Journal of Chemical Physics* 118 (15) (2003) 7127–7140.
- [12] Y. Zhang, D. M. Crothers, Statistical mechanics of sequence-dependent circular DNA and its application for DNA cyclization, *Biophysical Journal* 84 (1) (2003) 136–153.
- [13] R. Dickerson, M. Bansal, C. Calladine, S. Diekmann, W. Hunter, O. Kennard, E. von Kitzing, R. Lavery, H. Nelson, W. Olson, W. Saenger, Z. Shakked, H. Sklenar, D. Soumpasis, C.-S. Tung, A.-J. Wang, V. Zhurkin, Definitions and nomenclature of nucleic acid structure parameters, *Journal of Molecular Biology* 205 (4) (1989) 787–791.
- [14] M. A. El Hassan, C. R. Calladine, The assessment of the geometry of dinucleotide steps in double-helical DNA; a new local calculation scheme, *Journal of Molecular Biology* 251 (5) (1995) 648–664.
- [15] J. Langer, D. A. Singer, Lagrangian aspects of the kirchhoff elastic rod, *SIAM Review* 38 (1996) 605–618.
- [16] X.-J. Lu, W. K. Olson, 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures, *Nature Protocols* 3 (7) (2008) 1213–1227.



- [17] C. F. Gauss, Carl Friedrich Gauss Werke ... Herausgegeben von der K. Gesellschaft der Wissenschaften zu Göttingen, Göttingen, Gedruckt in der Dieterichschen Universitätsdruckerei (W.F. Kaestner), 1867.
- [18] J. H. White, W. R. Bauer, Calculation of the twist and the writhe for representative models of DNA, *Journal of Molecular Biology* 189 (2) (1986) 329–341.
- [19] W. Pohl, The self-linking number of a closed space curve, *Indiana Univ. Math. J.* 17 (1968) 975–985.
- [20] N. Clauvelin, W. K. Olson, I. Tobias, Characterization of the geometry and topology of DNA pictured as a discrete collection of atoms, *Journal of Chemical Theory and Computation* 8 (3) (2012) 1092–1107.
- [21] E. E. Zajac, Stability of two planar loop elasticas, *ASME J. Applied Mechanics* 29 (1962) 136–142.
- [22] A. Goriely, Twisted Elastic Rings and the Rediscoveries of Michell’s Instability, *Journal of Elasticity* 84 (3) (2006) 281–299.
- [23] S. Kim, E. Broströmer, D. Xing, J. Jin, S. Chong, H. Ge, S. Wang, C. Gu, L. Yang, Y. Q. Gao, X.-d. Su, Y. Sun, X. S. Xie, Probing allostery through DNA, *Science* 339 (6121) (2013) 816–819.
- [24] K. Kobryn, D. Z. Naigamwalla, G. Chaconas, Site-specific DNA binding and bending by the *Borrelia burgdorferi* Hbb protein, *Molecular Microbiology* 37 (1) (2000) 145–155.
- [25] S. Adhya, Multipartite genetic control elements: Communication by DNA loop, *Annual Review of Genetics* 23 (1) (1989) 227–250.
- [26] R. Schleif, DNA looping, *Annual Review of Biochemistry* 61 (1) (1992) 199–223.
- [27] X. Xu, H. Ge, C. Gu, Y. Q. Gao, S. S. Wang, B. J. R. Thio, J. T. Hynes, X. S. Xie, J. Cao, Modeling spatial correlation of DNA deformation: DNA allostery in protein binding, *The Journal of Physical Chemistry B* 117 (42) (2013) 13378–13387.
- [28] J. F. Burd, R. M. Wartell, J. B. Dodgson, R. D. Wells, Transmission of stability (telestability) in deoxyribonucleic acid. Physical and enzymatic studies on the duplex block polymer d(C15A15)-d(T15G15), *Journal of Biological Chemistry* 250 (13) (1975) 5109–13.
- [29] F. Touzain, M.-A. Petit, S. Schbath, M. E. Karoui, DNA motifs that sculpt the bacterial chromosome, *Nat Rev Micro* 9 (1) (2011) 15–26.
- [30] S. Bochkhanov, V. Bystritsky, ALGLIB, <http://www.alglib.net>.
- [31] G. Guennebaud, B. Jacob, et al., Eigen v3, <http://eigen.tuxfamily.org> (2010).

## Appendix A. Base-pair step geometry

We consider in this appendix and the following ones a collection of  $N$  rigid base pairs and for the  $i$ -th base pair we denote  $\underline{x}^i$  its origin and  $\mathbf{d}^i$  the matrix containing the axes of the base-pair frame organized as column vectors. The step parameters of the  $i$ -th step are denoted  $\underline{p}^i$  and the step dofs are denoted  $\underline{\phi}^i$ . The definitions of the step parameters are given in [11, 13, 14] (in particular, the first reference provides expressions for the angular step parameters in terms of the Euler angles describing the rotation between two successive base pairs).

Vector symbols are underlined and matrices are represented with bold symbols and the elements of a vector or a matrix are denoted with square brackets. Superscripted Greek letters ( $\alpha, \beta, \dots$ ) are used to denote vector, matrix, or tensor entries and range from 1 to 3. Superscripted Roman letters ( $i, j, \dots$ ), on the other hand, are used to index base pairs and base-pair steps. For example, the  $\alpha$ -th axis of base-pair frame  $\mathbf{d}^i$  is denoted  $\underline{d}_\alpha^i$ , and the  $\beta$ -th component of that vector is  $[\underline{d}_\alpha^i]_\beta = [\mathbf{d}^i]_{\alpha\beta}$ . We also use the Einstein summation notation on lower indices: for example, the scalar product of two vectors  $\underline{u}$  and  $\underline{v}$  is given by  $[\underline{u}]_\alpha [\underline{v}]_\alpha$ , that is,  $\underline{u} \cdot \underline{v} = [\underline{u}]_\alpha [\underline{v}]_\alpha = \sum_{\alpha=1}^3 [\underline{u}]_\alpha [\underline{v}]_\alpha$ .

### Appendix A.1. Step rotation

For a given base-pair step, we denote  $\mathbf{D}^i$  the matrix describing the rotation between the two successive base-pair frames  $\mathbf{d}^i$  and  $\mathbf{d}^{i+1}$ . This rotation matrix is referred to as the step rotation matrix and is defined by:

$$\mathbf{D}^i(\underline{\psi}^i) = \mathbf{d}^{i\top} \mathbf{d}^{i+1}. \quad (\text{A.1})$$

One of the main results about the geometry of a base-pair step is the relation between the changes in the angular step dofs and the changes in the relative orientation of the base-pair frames. We assume that the base-pair frames are orthonormal and it follows that an infinitesimal perturbation of a base-pair frame axis  $\underline{d}_\alpha^i$  can be written as  $\delta \underline{d}_\alpha^i = \underline{\chi}^i \times \underline{d}_\alpha^i$ . The changes in the step rotation matrix due to the perturbation of the two base-pair frames forming the step is therefore given by:

$$[\delta \mathbf{D}^i]_{\alpha\beta} = \delta \underline{\chi}^i \cdot \underline{d}_\beta^{i+1} \times \underline{d}_\alpha^i = -[\mathbf{D}^i]_{\gamma\beta} \epsilon_{\alpha\gamma\zeta} \underline{d}_\zeta^i \cdot \delta \underline{\chi}^i, \quad (\text{A.2})$$

where  $\delta \underline{\chi}^i = \underline{\chi}^{i+1} - \underline{\chi}^i$  and  $\epsilon_{\alpha\gamma\zeta}$  is the three-dimensional Levi-Civita symbol. We introduce the vector  $\delta \underline{\omega}^i = \mathbf{d}^{i\top} \delta \underline{\chi}^i$  which is the projection of  $\delta \underline{\chi}^i$  in the  $i$ -th base-pair frame. It follows that:

$$[(\delta \mathbf{D}^i) \mathbf{D}^{i\top}]_{\alpha\beta} = -\epsilon_{\alpha\beta\gamma} [\delta \underline{\omega}^i]_\gamma. \quad (\text{A.3})$$

The term on the left-hand side corresponds to a skew matrix because of the orthogonality of  $\mathbf{D}^i$  and we introduce the matrix  $\Xi^i$  such that:

$$[(\delta \mathbf{D}^i) \mathbf{D}^{i\top}]_{\alpha\beta} = \left[ \frac{\partial \mathbf{D}^i}{\partial \psi_\gamma^i} \mathbf{D}^{i\top} \right]_{\alpha\beta} \delta \psi_\gamma^i = -\epsilon_{\alpha\beta\zeta} [\Xi^i]_{\zeta\gamma} \psi_\gamma^i \quad (\text{A.4})$$

We finally obtain after contracting the Levi-Civita symbols:

$$\delta \underline{\omega}^i = \Xi^i \delta \underline{\psi}^i. \quad (\text{A.5})$$

This expression relates the relative changes in the base-pair frames to the variations of the angular step dofs. The matrix  $\Xi^i$  is invertible [11] and we introduce the matrix  $\Omega^i = \Xi^{i-1}$  to obtain:

$$\delta \underline{\psi}^i = \Omega^i \delta \underline{\omega}^i. \quad (\text{A.6})$$

The matrix  $\Omega^i$  can be obtained by direct calculation and its transpose is identical to the matrix  $\Gamma^n$  given in Eq. (A.1) of [11].

### Appendix A.2. Step frame

The step itself can be characterized by introducing a so-called step frame (referred to as the mid-step frame in [14]). This frame is located at  $(\underline{x}^i + \underline{x}^{i+1})/2$  and the matrix containing the step frame axes as columns, denoted  $\mathbf{s}^i$ , is given by:

$$\mathbf{s}^i = \mathbf{d}^i \mathbf{D}_s^i(\underline{\psi}^i), \quad (\text{A.7})$$

where  $\mathbf{D}_s^i$  is a rotation matrix that can be obtained directly from the angular step dofs (see [11] for more details).

### Appendix A.3. Translational step parameters

The step parameters  $\underline{\rho}^i$  are defined as the projection of the step joining vector  $\underline{r}^i = \underline{x}^{i+1} - \underline{x}^i$  on the step frame, that is, we have:

$$\underline{\rho}^i = \mathbf{s}^{i\top} \underline{r}^i. \quad (\text{A.8})$$

The definition given by Eq. (A.8) shows that the step parameters  $\underline{\rho}^i$  depend on the step dofs  $\underline{r}^i$  and on the angular step dofs  $\underline{\psi}^i$ . We introduce the matrix  $\mathbf{T}^i$  such that:

$$\delta \underline{\rho}^i = \mathbf{T}^i \delta \underline{r}^i, \quad (\text{A.9})$$

and we have:

$$[\mathbf{T}^i]_{\alpha\beta} = \frac{\partial \rho_\alpha^i}{\partial r_\beta^i} = [\mathbf{s}^{i\top}]_{\alpha\beta}. \quad (\text{A.10})$$

We also introduce the matrix  $\mathbf{R}^i$  such that:

$$\delta \underline{\rho}^i = \mathbf{R}^i \delta \underline{\psi}^i. \quad (\text{A.11})$$

We have:

$$[\mathbf{R}^i]_{\alpha\beta} = \frac{\partial \rho_\alpha^i}{\partial \psi_\beta^i} = \left[ \frac{\partial \mathbf{D}_s^{i\top}}{\partial \psi_\beta^i} \mathbf{D}_s^i \right]_{\alpha\gamma} \rho_\gamma^i, \quad (\text{A.12})$$

where the matrix appearing in the right-hand side is skew ( $\mathbf{D}_s^i$  is orthogonal). We denote  $\underline{\Lambda}_\beta^i$  its vector representations and we obtain:

$$[\mathbf{R}^i]_{\alpha\beta} = [\underline{\Lambda}_\beta^i \times \underline{\rho}^i]_\alpha. \quad (\text{A.13})$$

The vectors  $\underline{\Lambda}_\beta^i$  can be obtained by direct calculation and are identical to the ones given in Eqs (A.2-A.4) in [11].

## Appendix B. Base-pair collection geometry

We derive in this appendix the results related to the Jacobian matrix  $\mathbf{J}_\Phi$ . This Jacobian matrix is defined as (see Eq. (5)):

$$\mathbf{J}_\Phi = \frac{\partial \underline{P}}{\partial \Phi}, \quad (\text{B.1})$$

where  $\underline{P}$  denotes the set of all base-pair step parameters and  $\Phi$  the set of all step dofs.

*Appendix B.1. Rotations within a base-pair collection*

We denote  $\mathcal{D}^{(i,j)}$  the matrix describing the rotation between the base-pair frames  $\mathbf{d}^i$  and  $\mathbf{d}^j$  ( $i < j$ ). This rotation matrix is defined as:

$$\mathcal{D}^{(i,j)} = \mathbf{d}^{i\top} \mathbf{d}^j = \prod_{k=i}^{j-1} \mathbf{D}^k(\underline{\psi}^k), \quad (\text{B.2})$$

where  $\mathbf{D}^k(\underline{\psi}^k)$  is the step rotation matrix (see Eq. (A.1)).

We want to calculate the derivatives of this product of rotation matrices with respect to the angular step dofs  $\underline{\psi}^k$  ( $i \leq k < j$ ). We have:

$$\frac{\partial \mathcal{D}^{(i,j)}}{\partial \psi_\alpha^k} = \mathbf{D}^i \dots \mathbf{D}^{k-1} \frac{\partial \mathbf{D}^k}{\partial \psi_\alpha^k} \mathbf{D}^{k+1} \dots \mathbf{D}^{j-1}, \quad (\text{B.3})$$

which can be rewritten as:

$$\frac{\partial \mathcal{D}^{(i,j)}}{\partial \psi_\alpha^k} = \mathbf{d}^{i\top} \mathbf{d}^k \left( \frac{\partial \mathbf{D}^k}{\partial \psi_\alpha^k} \mathbf{D}^{k\top} \right) \mathbf{d}^{k\top} \mathbf{d}^j. \quad (\text{B.4})$$

The matrix in parentheses on the right-hand side is a skew matrix (see Eq. (A.4)) and we introduce the notation  $(\partial \mathbf{D}^k / \partial \psi_\alpha^k) \mathbf{D}^{k\top} = \mathbf{P}_\alpha^k$  such that:

$$\frac{\partial \mathcal{D}^{(i,j)}}{\partial \psi_\alpha^k} = \mathbf{d}^{i\top} \mathbf{d}^k \mathbf{P}_\alpha^k \mathbf{d}^{k\top} \mathbf{d}^j = \mathcal{D}^{(i,k)} \mathbf{P}_\alpha^k \mathcal{D}^{(k,j)}, \quad (\text{B.5})$$

and the transpose is given by:

$$\left( \frac{\partial \mathcal{D}^{(i,j)}}{\partial \psi_\alpha^k} \right)^\top = \mathcal{D}^{(k,j)\top} \mathbf{P}_\alpha^{k\top} \mathcal{D}^{(i,k)\top} = -\mathcal{D}^{(k,j)\top} \mathbf{P}_\alpha^k \mathcal{D}^{(i,k)\top}, \quad (\text{B.6})$$

where we used the property  $\mathbf{P}_\alpha^{k\top} = -\mathbf{P}_\alpha^k$ .

We can use this result to calculate the derivatives of a base-pair frame with respect to the angular step dofs of the preceding steps. The  $i$ -th base-pair frame in the collection is given by:

$$\mathbf{d}^i = \mathbf{d}^1 \mathcal{D}^{(1,i)}. \quad (\text{B.7})$$

It follows from Eq. (B.5) that:

$$\frac{\partial \mathbf{d}^i}{\partial \psi_\alpha^k} = \left( \mathbf{d}^k \mathbf{P}_\alpha^k \mathbf{d}^{k\top} \right) \mathbf{d}^i \quad (\text{B.8})$$

The matrix multiplying  $\mathbf{d}^i$  on the right-hand side is a skew matrix and can therefore be represented by a vector. This vector expresses the infinitesimal rotation due to changes in the  $\underline{\psi}^k$ . We define the vector  $\underline{S}_\alpha^k$  as:

$$\left[ \underline{S}_\alpha^k \right]_\beta = -\frac{1}{2} \epsilon_{\gamma\zeta\beta} \left[ \mathbf{d}^k \mathbf{P}_\alpha^k \mathbf{d}^{k\top} \right]_{\gamma\zeta}. \quad (\text{B.9})$$

This expression can be rewritten in a more concise form with the help of Eq. (A.4) ( $\mathbf{P}_\alpha^k$  is the skew matrix obtained from the  $\alpha$ -th column of  $\Xi^k$ ):

$$\underline{S}_\alpha^k = \mathbf{d}^j \Xi_\alpha^k, \quad (\text{B.10})$$

where  $\Xi_\alpha^k$  is the  $\alpha$ -th column of  $\Xi^k$ . In other words,  $\underline{S}_\alpha^k$  is the expression of the vector  $\Xi_\alpha^k$  in the global reference frame and the vector  $\underline{S}_\alpha^k$  describes the infinitesimal rotation associated with a change in the  $\alpha$ -th angular step dof of the  $k$ -th step. We finally obtain:

$$\frac{\partial}{\partial \psi_\alpha^k} [\mathbf{d}^k]_{\gamma\zeta} = \frac{\partial}{\partial \psi_\alpha^k} [\underline{d}_\zeta^k]_\gamma = [\underline{S}_\alpha^k \times \underline{d}_\zeta^k]_\gamma. \quad (\text{B.11})$$

### Appendix B.2. Jacobian matrix $\mathbf{J}_\Phi$

We already have two of the three contributions to the Jacobian matrix of the collection: the matrices  $\mathbf{T}^i$  and  $\mathbf{R}^i$  (see Eq. (A.10) and Eq. (A.13)) express the dependence of the  $\underline{\rho}^i$  on the step dofs  $\underline{\psi}^i$  and  $\underline{r}^i$ . The definition of the  $\underline{\rho}^i$  (Eq. (A.8)) shows that there is also a dependence on the angular dofs from the preceding steps in the collection. We have from Eq. (A.8):

$$\frac{\partial \rho_\alpha^i}{\partial \psi_\beta^j} = \left[ \mathbf{D}_s^i{}^\top \left( \frac{\partial \mathbf{d}^i}{\partial \psi_\beta^j} \right)^\top \right]_{\alpha\gamma} r_\gamma^i, \text{ where } j < i \quad (\text{B.12})$$

which, with the help of Eq. (B.11), leads to:

$$\frac{\partial \rho_\alpha^i}{\partial \psi_\beta^j} = - \left[ \left( \mathbf{s}^{i\top} \underline{\Sigma}_\beta^j \right) \times \underline{\rho}^i \right]_\alpha = [\mathbf{U}^{j,i}]_{\alpha\beta}. \quad (\text{B.13})$$

We can now write the full variation of the step parameters  $\underline{p}^i$  in terms of the step dofs  $\underline{\phi}^k$  ( $1 \leq k \leq i$ ) as:

$$\delta \underline{p}^i = \left[ \begin{array}{c|c} \mathbf{I}_3 & \mathbf{0} \\ \hline \mathbf{R}^i & \mathbf{T}^i \end{array} \right] \delta \underline{\phi}^i + \sum_{j=1}^{i-1} \left[ \begin{array}{c|c} \mathbf{0} & \mathbf{0} \\ \hline \mathbf{U}^{j,i} & \mathbf{0} \end{array} \right] \delta \underline{\phi}^j, \quad (\text{B.14})$$

where  $\mathbf{R}^i$  is defined in Eq. (A.13),  $\mathbf{T}^i$  in Eq. (A.10), and  $\mathbf{U}^{j,i}$  in Eq. (B.13). The matrix  $\mathbf{J}_\Phi$  is directly obtained from this expression and its structure is shown in Fig. B.9.

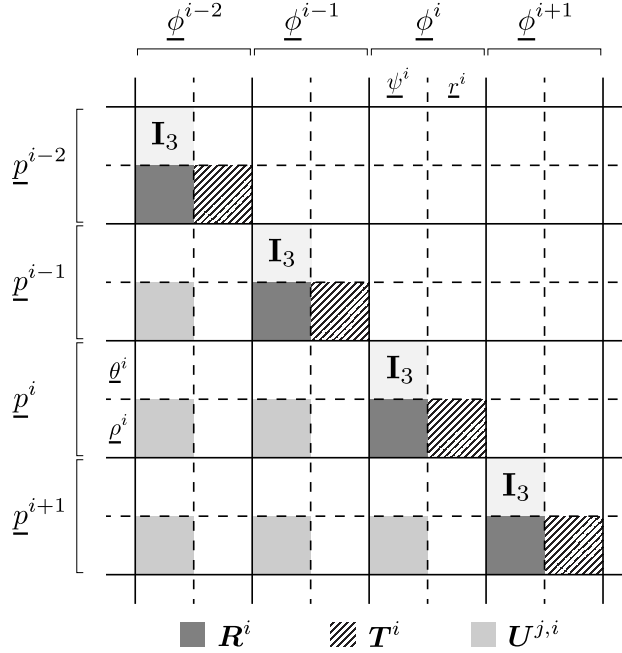


Figure B.9: Structure of the matrix  $\mathbf{J}_\Phi$  for four *central* steps  $i-2, \dots, i+1$ . The vectors  $\underline{p}^i$  denote the step parameters while the vectors  $\underline{\phi}^i$  denote the step dofs. The solid lines delimit the  $6 \times 6$  base-pair step block matrices, while the dashed lines separate the angular variables from the translational variables ( $\underline{\theta}^i$  and  $\underline{\rho}^i$  for the step parameters, and  $\underline{\psi}^i$  and  $\underline{r}^i$  for the step dofs). The white blocks represent null entries.

### Appendix C. End conditions

We provide in this appendix the details for the calculation of the Jacobian matrix  $\hat{\mathbf{J}}$ . This Jacobian matrix relates the set of *independent* step dofs to the complete set of step dofs. The number of independent step dofs depends on the details of the boundary conditions. We focus here on an imposed end-to-end vector and an imposed end-to-end rotation but other types of boundary conditions can be treated along the same lines.

The imposed end-to-end vector condition is given by Eq. (10) which reads:

$$\delta \underline{\mathbf{r}}^{N-1} = - \sum_{i=1}^{N-2} \delta \underline{\mathbf{r}}^i. \quad (\text{C.1})$$

In other words, the change in the joining vector of the last step is completely determined by the changes in the joining vectors of the other steps.

The imposed end-to-end rotation condition corresponds to (see Eq. (12)):

$$\delta \mathcal{D}^{(1,N)} = \delta \left( \prod_{i=1}^{N-1} \mathbf{D}^i(\underline{\psi}^i) \right) = \sum_{i=1}^{N-1} \frac{\partial \mathcal{D}^{(1,N)}}{\partial \psi_\alpha^i} \delta \psi_\alpha^i = \mathbf{0}. \quad (\text{C.2})$$

This condition can be written as:

$$\frac{\partial \mathcal{D}^{(1,N)}}{\partial \psi_\alpha^{N-1}} \delta \psi_\alpha^{N-1} = - \sum_{i=1}^{N-2} \frac{\partial \mathcal{D}^{(1,N)}}{\partial \psi_\alpha^i} \delta \psi_\alpha^i, \quad (\text{C.3})$$

and with the help of Eq. (B.5) we obtain:

$$\mathbf{P}_\alpha^{N-1} \delta \psi_\alpha^{N-1} = - \mathbf{d}^{N-1 \top} \left( \sum_{i=1}^{N-2} \mathbf{d}^i \mathbf{P}_\alpha^i \mathbf{d}^{i \top} \right) \mathbf{d}^{N-1} \delta \psi_\alpha^{N-1}. \quad (\text{C.4})$$

The right-hand side corresponds to a sum of skew matrices and we therefore introduce  $\mathbf{Q}^{i,N-1}$  as:

$$\left[ \mathbf{d}^{N-1 \top} \mathbf{d}^i \mathbf{P}_\alpha^i \mathbf{d}^{i \top} \mathbf{d}^{N-1} \right]_{\beta\gamma} \delta \psi_\alpha^i = - \epsilon_{\beta\gamma\zeta} [\mathbf{Q}^{i,N-1}]_{\zeta\alpha} \delta \psi_\alpha^i. \quad (\text{C.5})$$

Using Eq. (A.4) we obtain after contracting the Levi-Civita symbols:

$$[\Xi^{N-1}]_{\zeta\alpha} \delta \psi_\alpha^{N-1} = - \sum_{i=1}^{N-2} [\mathbf{Q}^{i,N-1}]_{\zeta\alpha} \delta \psi_\alpha^i. \quad (\text{C.6})$$

We can invert the left-hand side (see Eq. (A.6)) to finally obtain:

$$\delta \underline{\psi}^{N-1} = - \sum_{i=1}^{N-2} \Omega^{N-1} \mathbf{Q}^{i,N-1} \delta \underline{\psi}^i = - \sum_{i=1}^{N-2} \mathbf{K}^i \delta \underline{\psi}^i. \quad (\text{C.7})$$

Note that the matrix  $\mathbf{Q}^{i,N-1}$  can be expressed in terms of the vectors  $\underline{\mathcal{S}}_\beta^i$ :

$$[\mathbf{Q}^{i,N-1}]_{\alpha\beta} = \underline{\mathcal{S}}_\alpha^{N-1} \cdot \underline{\mathcal{S}}_\beta^i. \quad (\text{C.8})$$

In other words, the matrix  $\mathbf{Q}^{i,N-1}$  is obtained by projecting the vectors  $\underline{\mathcal{S}}_\beta^i$  in the base-pair frame  $\mathbf{d}^{N-1}$  and, hence, expresses the infinitesimal rotation originating from the changes in the angular step dofs  $\underline{\psi}^i$  in the base-pair frame  $\mathbf{d}^{N-1}$ .

The results obtained in Eq. (10) and Eq. (C.7) show that:

$$\delta \underline{\phi}^{N-1} = - \sum_{i=1}^{N-2} \left[ \begin{array}{c|c} \mathbf{K}^i & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{I}_3 \end{array} \right] \delta \underline{\phi}^i = - \sum_{i=1}^{N-2} \mathbf{B}^i \delta \underline{\phi}^i. \quad (\text{C.9})$$

This shows that the imposed end-to-end vector and rotation conditions reduce the number of independent step dofs (the step dofs of the last step can be expressed in terms of the step dofs of all the other steps). That is, the set of independent step dofs, denoted  $\hat{\underline{\Phi}}$ , corresponds to  $\hat{\underline{\Phi}} = \{\underline{\phi}^i\}_{i=1,\dots,N-2}$ . The Jacobian matrix  $\hat{\mathbf{J}}$ , defined as  $\hat{\mathbf{J}} = \partial \underline{\Phi} / \partial \hat{\underline{\Phi}}$ , is sparse and its structure is shown in Fig. C.10.

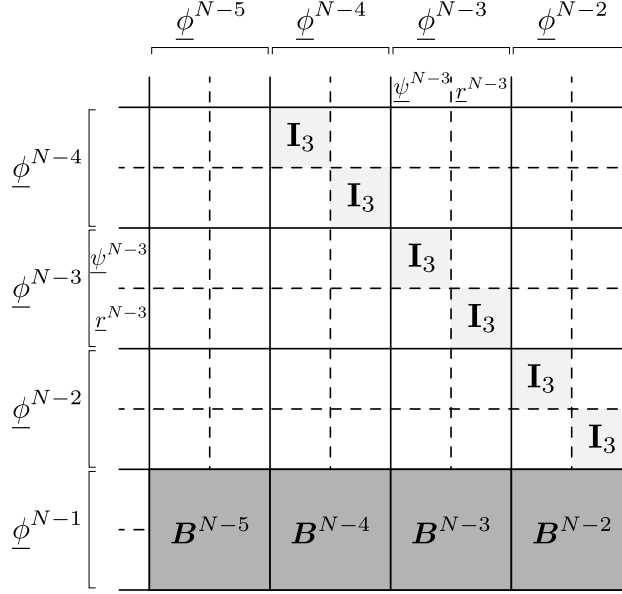


Figure C.10: Structure of the matrix  $\hat{\mathbf{J}}$  for imposed end-to-end vector and rotation for the last four steps of a collection of base pairs. The vectors  $\underline{\phi}^i$  denote the step dofs. The columns of this matrix are related to the independent step dofs and, hence, the last column is for step  $(N-2)$  as the  $(N-1)$ -th step is not independent due to the boundary conditions. The solid lines delimit the  $6 \times 6$  base-pair step block matrices, while the dashed lines separate the angular variables  $\underline{\psi}^i$  from the translational variables  $\underline{r}^i$ . The white blocks represent null entries.

## Appendix D. Frozen steps

We focus in this appendix on the calculation of the matrix  $\tilde{\mathbf{J}}$  (see Eq. (20)).

We consider that the  $k$ -th step in the collection of base pairs is frozen, that is, its step parameters  $\underline{p}^k$  are imposed and constant. It follows directly that:

$$\delta \underline{\psi}^k = \underline{0}. \quad (\text{D.1})$$

The condition on the variation of the translational step dofs,  $\delta \underline{r}^k$ , is obtained from the expression:

$$\delta \underline{\rho}^k = \frac{\partial \rho_\alpha^k}{\partial r_\beta^k} \delta r_\beta^k + \sum_{\ell=1}^{k-1} \frac{\partial \rho_\alpha^k}{\partial \psi_\beta^\ell} \delta \psi_\beta^\ell = \underline{0}. \quad (\text{D.2})$$

It follows from Eq. (A.10) and Eq. (B.13):

$$\delta r_\beta^k = \sum_{\ell=1}^{k-1} \left[ \underline{S}_\gamma^\ell \times \underline{r}^k \right]_\beta \delta \psi_\gamma^\ell = \sum_{\ell=1}^{k-1} [\mathbf{W}^{\ell,k}]_{\beta\gamma} \delta \psi_\gamma^\ell. \quad (\text{D.3})$$

This result simply mean that the vector  $\underline{r}^k$  only changes in orientation due to the changes in the angular step dofs of the preceding steps.

We introduce the matrix  $\mathbf{C}^{\ell,k}$  as:

$$\delta \underline{\phi}^k = \sum_{\ell=1}^{k-1} \mathbf{C}^{\ell,k} \delta \underline{\phi}^\ell = \sum_{\ell=1}^{k-1} \left[ \begin{array}{c|c} \mathbf{0} & \mathbf{0} \\ \hline \mathbf{W}^{\ell,k} & \mathbf{0} \end{array} \right] \delta \underline{\phi}^\ell. \quad (\text{D.4})$$

The Jacobian matrix  $\tilde{\mathbf{J}}$  is readily determined from this result and its structure is shown in Fig. D.11.

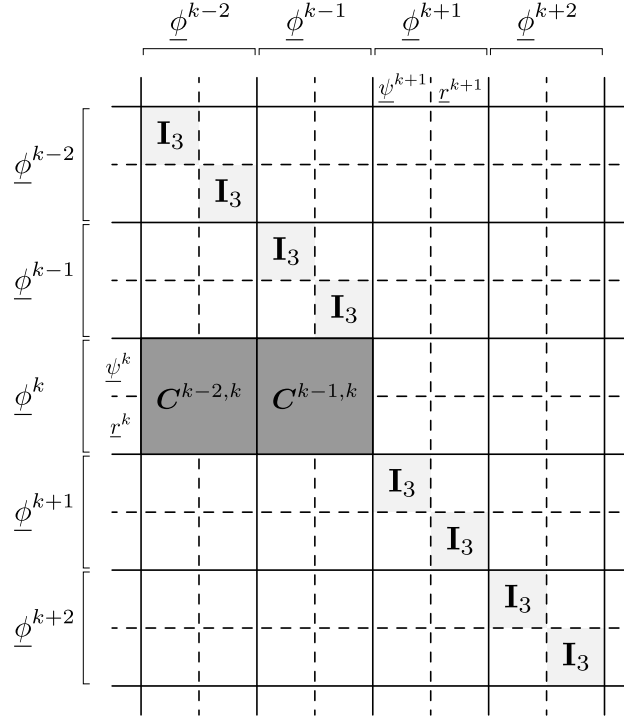


Figure D.11: Structure of the matrix  $\tilde{\mathbf{J}}$  for a base-pair collection in which the  $k$ -th step is frozen. The solid lines delimit base-pair step block matrices, while the dashed lines separate the angular variables  $\underline{\psi}^i$  from the translational variables  $\underline{r}^i$ . The white blocks represent null entries.

## Appendix E. Zajac instability for DNA minicircles

The Zajac instability [21, 22] is related to the fact that a twisted elastic ring becomes unstable if the twist density within the ring exceeds a certain value. It is straightforward to transpose this result to our DNA base-pair model (in fact, our mesoscale model together with our ideal force field can be seen as a discretized Kirchhoff elastic rod model).

For a DNA fragment of  $N$  bp in its rest state, the intrinsic total twist  $\overline{\text{Tw}}$  is given by:

$$\overline{\text{Tw}} = \frac{N * \bar{\theta}_3}{360} = \frac{N}{10.5}, \quad (\text{E.1})$$

where  $\bar{\theta}_3 = 360/10.5 = 34.2857$  deg is the intrinsic twist step parameter value. We introduce the excess of twist  $\Delta \text{Tw} = \text{Tw} - \overline{\text{Tw}}$  and the stability criterion is written as:

$$|\Delta \text{Tw}| < \frac{\sqrt{3}}{2\Gamma}, \quad (\text{E.2})$$



where  $\Gamma$  is the ratio of the bending and twisting stiffnesses (for our ideal for field we have  $\Gamma = 0.715$ ). Since we only consider the case of a planar minicircle we have  $\text{Tw} = \text{Lk} = \Delta\text{Lk} + \text{Lk}^0$  and we recall that  $\text{Lk}^0 = [N/10.5]$  (where the brackets stand for the nearest integer operator). It follows that the criterion can be written as:

$$-\frac{\sqrt{3}}{2\Gamma} - \left( \left[ \frac{N}{10.5} \right] - \frac{N}{10.5} \right) < \Delta\text{Lk} < \frac{\sqrt{3}}{2\Gamma} - \left( \left[ \frac{N}{10.5} \right] - \frac{N}{10.5} \right). \quad (\text{E.3})$$

These conditions with the fact that  $\Delta\text{Lk}$  is an integer leads to the following set of solutions:

$$\Delta\text{Lk} = \{-1, 0, +1\}. \quad (\text{E.4})$$

## Appendix F. Implementation details

We use the C++ ALGLIB library [30] to implement our minimization method and rely on the L-BFGS algorithm. We also use the C++ Eigen library [31] to handle the various matrix computations and manipulations. Our minimization software is available upon request.